Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2000 Proceedings

Americas Conference on Information Systems (AMCIS)

2000

Making Theories More Accessible: A Proposal for a Theory Markup Language

Marc N. Haines
The University of Georgia, mhaines@uga.edu

Making Theories More Accessible: A Proposal for a Theory Markup Language

Marc N. Haines, Department of Management Information Systems, Terry College of Business, The University of Georgia, mhaines@uga.edu

Abstract

This paper proposes a markup language based on XML for the domain of scientific theory, called theory markup language (TML). The author argues that this markup language allowing structured representations of theories, has the potential to make the search for and analysis of relevant theory more effective. In the first part of the paper the role of theory and its basic elements are discussed. Then a prototype of a document type definition for TML is presented. Finally issues beyond the technical definition of the markup language are addressed; including issues that are relevant for making a theory repository based on TML documents a useful resource for researchers in MIS and other disciplines.

Introduction

According to Davis and Cosenza (Davis, 1988), the establishment of theories is the primary goal of science, because they are the means by which we explain and predict phenomena of interest. Theories carry the essential body of knowledge of a scientific field. The purpose of a theoretical statement is to parsimoniously organize and clearly communicate (Bacharach, 1989) this knowledge. A theory can be defined as a systematic explanation for the observations that relate to a particular aspects of life (Babbie, 1998) or more formally as a statement of relations among concepts within a set of boundary assumptions and constraints (Bacharach, 1989). Other definitions offer slight variations (e.g. Kerlinger, 1986), but the basic elements of a theory appear to be constructs (synonym with concepts (Kerlinger, 1986)), variables, and relationships (Dubin, 1969; Schwab, 1980). In addition to these basic elements the context in which a theory is applied with its set of boundary assumptions and constraints needs to be considered.

The manifestations of theories can be found in scientific publications. Theories can be expressed in several different ways. This includes natural language descriptions, graphical representations, and mathematical formulas. For example the theory of reasoned action (TRA) is represented in all three ways in (Ajzen and Fishbein, 1972). A key principle of science is that research builds on the existing knowledge in a scientific field (Kerlinger, 1986). Since theories carry the existing body of knowledge a researcher needs to be aware of the relevant existing theories and understand their essence when starting out on a new research endeavor.

Currently, a researcher typically engages in a literature search to find theories or constructs that are relevant to the research. While scientific literature becomes more and more accessible due to electronic media and online databases, there are still significant limitations. In most cases the researcher is limited to search for a combination of keywords that may relate to any part of a document. A direct and explicit search for theories, particular constructs, or other basic elements of a theory, relationships between theories, and applications of theory is hardly possible. A researcher has to scan through many, often irrelevant, documents and extract the relevant theoretical knowledge. This process lacks efficiency and also bears the potential for omitting relevant theories.

This article proposes a structure for documents describing theory. This structure has the potential to make the process of identifying and analyzing relevant theory easier and more reliable. By using the general extensible markup language (XML) (Bray et al., 1998) the author develops a specific markup language for the purpose of describing scientific theories called TML (theory markup language). The development of this markup language is seen as a first step towards a shared theoretical knowledge repository for the research community in MIS and other disciplines. In addition to TML, other mechanisms and tools to create, publish, retrieve, and process TML documents will have to be developed, implemented, and adopted by the research community to make this a useful resource to researchers.

The Basic Elements of Theory

Based on a review of theory related articles (Bacharach, 1989, Dubin, 1969, Kerlinger, 1986) and MIS related articles (i.e. Davis et al., 1989) the author identified the following key elements of theory shown in table 1.

Table 1. Elements of Theory

Element Name	Element Description	TML Element Name
Theory	Description of the theory and its context	<THEORY>
Construct	Definition of constructs proposed in the theory	<CONSTRUCT>
Relationship	Definition of the relationships among constructs proposed in the theory	<REL>
Theory Application	The application of a theory in a particular study	<THEOAPP>

Element Name	Element Description	TML Element Name
Instrument	Definition of the instrument used in the application of the theory	<INSTRUMENT>
Measure	Definition of a measure relating to a certain construct in the theory	<MEASURE>
Variable	Definition of a variable which is part of a measure	<VAR>
Hypothesis	Description of hypotheses relating to certain relationships in the theory	<HYPO>
Results	Outcomes for instrument (global), measures, and variables in the instrument	<RESULTLIST>

A theory consists of a set of constructs and a set of relationships among those constructs. A theory can be applied in the context of a particular study. For each application, a set of hypotheses may be formulated. An application of a theory often involves the use of a particular instrument to measure the constructs proposed in the theory. A specific instrument associates a measure with each construct. Each measure consists of a set of variables. A study will yield several results related to the hypotheses, measures, variables, and the instrument.

While it is important to adequately describe a theory and its application in a particular study, it is at least as important to capture the evolution of a theory. Theories are often derived from other theories by combination, modification, or extension. Theories may also be seen as competitors with other theories. The theory of reasoned action (TRA) (Ajzen and Fishbein, 1972) and its extensions such as the theory of planned behavior (TPB) (Ajzen, 1991) are an example. By capturing these relationships among theories it is possible for a researcher to effectively search for related theories, their context of use, and outcomes in past studies.

An effective process of identifying relevant theory and its applications is especially important in a research discipline like management information systems, which borrows theories from many other disciplines.

TML Document Type Definition

The document type definition (DTD) defines the structure of a TML document. Only documents adhering to this DTD are considered valid TML documents. A prototype for a document type definition for TML is described in (Haines, 2000).

Each key element of theory listed in the table 1 has a corresponding tag in the theory markup language (e.g. relationship and <REL>). The attributes of a tag and the information between the tags carry the information for a theory element. A tag can have several nested subordinate tags as shown in figure 1.

Figure 1. DTD Example - Theory

```
<!ELEMENT TML (TMLHEAD, (THEORY | THEORYAPP |
(THEORY, THEORYAPP)))>

<!ELEMENT THEORY (THEOHEAD, THEOREFLIST?,
LITREFLIST?, CONSTRUCTLIST, RELLIST)>
```

For example, between a pair of theory tags, there has to be exactly a pair of theory head tags (THEOHEAD) with its contents, possibly a list of references to other theories (THEOREFLIST), possibly a list of references to relevant literature (LITREFLIST), exactly one list of constructs (CONSTRUCTLIST), and exactly one list of relationships (RELLIST).

Figure 2. DTD Example - Theory Reference

```
<!ELEMENT THEOREFLIST (THEOREF+)>

<!ELEMENT THEOREF (ID, LABEL,)>

<!ATTLIST THEOREF
  REFTYPE (internal | external) "internal"
  URI (#PCDATA) #implied
  USE (extends|modifies|competes) "modifies">
```

If a theory is derived from another theory, references to the parent theories can be included (see figure 2). Figure 3 shows an example of how TRA can be referenced in a TML document describing Theory of Planned Behavior.

Figure 3. TML Example - Theory of Planned Behavior

```
<TML>
…
<THEORY>
  <THEOEHAD>
    <ID>t002</ID>
    <LABEL>Theory of Planned Behavior</LABEL>
    <THEOABBREV>TPB</THEOABBREV>
  </THEOHEAD>
  < THEREFLIST>
    <THEOREF REFYPE="external"
                 URI="www.tml.org/tml/t001_tra.xml"
                 USE="extends">
      <ID>t001</ID>
      <LABEL>Theory of Reasoned Action</LABEL>
    </THEOREF>
  </THEOREFLIST>
          …
</THEORY>
</TML>
```

A key consideration in the development of TML is the balance between structure and flexibility. While clear structure enhances the ability to search for specific elements and values, TML has to be also flexible enough to accommodate a variety of theory configurations and unstructured annotations by the researcher.

Flexibility is further necessary to accommodate different types of research approaches, such as positivist or interpretivist research, that employ different ways of describing theory. It is important to consider the different ways in which theories are represented. Representations of positivist research tend to be rather structured and many elements and results can be described in the form of numbers, formulas, or other formal representations. Interpretivist research on the other hand tends to be less structured and elements and results are mostly described in natural language representations. Several of the TML elements described below will be more useful for the description of positivist research. Nonetheless, TML also includes free form text descriptions, which can be used to capture theories developed and described in interpretivist research. While some key TML elements are mandatory, there are many others that are optional and may only be applied if appropriate.

TML Example: Theory of Reasoned Action

The theory of reasoned action (TRA) (Ajzen and Fishbein, 1972) was chosen to demonstrate, how a theory can be represented in TML. The complete example at is described in (Haines, 2000). TRA has been widely used by MIS researchers to explain computer user behavior. Researchers seeking to improve the predictive power of the theory or model fit have introduced modifications of the original theory, such as the technology acceptance model (TAM) (Davis, 1986), the Miniard and Cohen Model (MCM) (Miniard, 1983), and the Theory of Planned Behavior (TPB) (Ajzen, 1991). Others have examined and compared those theories (Netemeyer and Bearden, 1992; Sheppard, 1988).

A researcher who, for example, is interested in studying computer user behavior would not only have to be aware of TRA but also it's the extensions and modifications. Currently a researcher can, for instance, search a reverse index to determine all the articles that cite a particular TRA article. The result of this search will most likely have a list of relevant and non-relevant articles and may miss some TRA related articles that do not reference this article directly. By reading each of the articles found in the search the researcher can then determine the relevant ones. The researcher may then decide that the information in the TAM article by Davis (Davis et al., 1989) and the TPB article by Chang (Chang, 1998) is relevant. It is then necessary to conduct another iteration of the reverse index search for those articles to explore further related articles and theories. If the information were available in TML format along with the appropriate search tools a representation of the related theories and a list of the articles referring to these theories could be generated automatically.

The search can also be more specific, due to the structure provided by TML. The focus can be on a specific construct (e.g. Behavioral Intention) and its measurements. The researcher can be provided with an overview of the relevant theories as well as detailed information about the constructs, instruments, and outcomes of a study. A defined structure may also facilitate automatic translations from and to other languages.

Technical, Organizational, and Political Challenges

Several technical, but more important organizational and political challenges have to be overcome to make a knowledge repository based on TML documents a useful resource. Technical issues are the development of a sound TML DTD, tools for creating and viewing TML documents, as well as establishing a search engine for TML documents. A search engine for TML documents should be capable of exploiting the semantic structure given in a TML document. This capability may be combined with other advanced search tech niques, including the use of case-based reasoning or artificial neural networks.

Probably the bigger challenges are on the organizational and political side. This includes the issues of broad adoption of TML, ownership of TML documents, and quality control of TML documents. A theory knowledge repository is only useful if it captures a critical mass of the available knowledge. Researchers have to be encouraged to create and publish TML documents related to their research. A possibility would be that journals and conferences require a corresponding TML document for each submission. Already existing theories would require a different approach. There is a potential for disagreement about the right representation of a theory and its applications in TML, in particular for theories where the original creator is unable to create his or her TML document. Alternative TML representations may coexist. To ensure the quality of TML documents a journal can have its repository of approved TML documents, which can be indexed in a global TML search engine. A researcher should then be able to decide which journals or journal categories she or he wants to include in a search. These are just a couple of suggestions for some of the issues that need to be resolved. Further investigation and feedback from the research community is necessary to evaluate the ideas presented in this article.

While TML documents should help a researcher to identify relevant theory, it is not intended to keep the researcher from reading the original in-depth articles to fully understand a theory and the related studies.

Conclusion and Outlook

This paper is seen less as a research paper than a proposal for a mechanism to make the research process more effective. The benefits envisioned are a more efficient and complete identification of relevant theory, including cross-disciplinary use of theories, and overcoming language barriers. Meta studies on theories and instruments (i.e. Sheppard, 1988) may also be supported using searches on TML documents.

For several domains (i.e. MathML for Mathematics) specific markup languages have already been introduced. Several markup languages for general business transactions (e.g. Microsoft BizTalk) are currently being developed, making XML a key technology in e-commerce transactions. While the author envisions various benefits from developing a markup language for scientific theories it is also clear that some difficult challenges have to be overcome to make TML a language of a broad knowledge repository for the scientific domain.

References

Ajzen, I. "The Theory of Planned Behavior," *Organizational Behavior & Human Decision Processes* (50), 1991, pp. 179-211.

Ajzen, I. and Fishbein, M. "Attitudes and Normative Beliefs as Factors Influencing Behavioral Intention," *Journal of Personality and Social Psychology* (21), 1972.

Babbie, E. *The Practice of Social Research*, Belmont, CA, Wadsworth Pub. Co., 8th edition, 1998.

Bacharach, S. B. "Organizational Theories: Some Criteria for Evaluation," *Academy of Management Review* (14:4), 1989, pp. 496-515.

Bray, T., Paoli J., Sperberg-McQueen C. M. "Extensible Markup Language (XML) 1.0 Specification," www.w3.org/TR/REC-xml, (Current February 10, 1998).

Chang, M. K. "Predicting Unethical Behavior: A Comparison of the Theory of Reasoned Action and the Theory of Planned Behavior," *Journal of Business Ethics* (17), 1998, pp. 1825-1834.

Davis, D., Consenza, Cosenza, R. M. *Business Research for Decision Making*, Boston, PWS-Kent Pub. Co., 1988.

Davis, F. D. *A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results*, Sloan School of Management, Massachusetts Institute of Technology, 1986.

Davis, F. D., et al. "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models," *Management Science* (35:8), 1989, pp. 982-1003.
Dubin, R. *Theory Building*. New York, Free Press, 1969.

Haines, M. N. "TML: An Introduction and Definition," www.marchaines.com/tml, (Current May 05, 2000).

Kerlinger, F. N. *Foundations of behavioral Research*, New York, Holt, Rinehart and Winston, 3rd edition, 1986.

Miniard, P. W., Cohen, Joel B. "Modeling Personal and Normative Influences on Behavior," *Journal of Consumer Research* (10), 1983, pp. 169-180.

Netemeyer, R. G. and Bearden, W.O. "A comparative Analysis of Two Models of Behavioral Intention." *Journal of the Academy of Marketing Science* (20:1), 1992, pp. 40-59.

Schwab, D. P. "Construct validity in organizational behavior" in Staw, B. M., Cummings, L. L., *Research in organizational behavior*, Greenwich, CT, JAI Press, 1980, pp. 3-43.

Sheppard, B. H., Hartwick, J., Warshaw, P. R. "The Theory of Reasoned Action: A Meta-Analysis of Past Research with Recommendations for Modifications and Future Research," *Journal of Consumer Research* (15), 1988, pp.325-343.